PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Complaint risk among mental health practitioners compared with physical health practitioners: A retrospective cohort study of complaints to health regulators in Australia
AUTHORS	Veness, Benjamin; Tibble, Holly; Grenyer, Brin; Morris, Jennifer; Spittal, Matthew; Nash, Louise; Studdert, David; Bismark, Marie

VERSION 1 – REVIEW

REVIEWER	Douglas Olsen Michigan State University, USA
REVIEW RETURNED	12-Apr-2019

GENERAL COMMENTS	Thank you for allowing me to review your work. It is a well done and useful study. It confirms prior literature and experience regarding the increased risk of mental health clinicians for complaints and boundary problems in a large sample.

REVIEWER	Jason Scott, Senior Lecturer Northumbria University, UK
REVIEW RETURNED	16-Apr-2019

GENERAL COMMENTS	This manuscript presents the results of a study on an interesting topic that uses established methods the authors have developed and used previously. However, I have a number of issues with the paper as it is currently presented, including with the methods used. 1. Abstract: the results section of the abstract is missing confidence intervals and p values for the reported statistics. These are required according to the journal guidelines. 2. Introduction: Whilst the introduction addresses differences in complaints relating to mental and physical health practitioners, it is lacking in a critical analysis of the literature surrounding complaints in healthcare and their implications for practice and policy. For instance, I would expect to see discussion surrounding the role that complaints have within the healthcare sector and for service improvement (linked to the previous comment about being 'informative'). There is a debate around whether complaints are a valid source of data for service improvement (including how they compare to other data sources), or whether they are treated as a litigation-avoidance exercise. 3. Introduction: In the hypothesis, the use of the word 'informative' is fairly meaningless. There needs to be an explanation as to why it could be informative, and to whom it would be informative for. This ties into the previous issue. 4. Methods: two reviewers independently coded 149 categories into 16 complaint issues, with differences resolved by consensus. I

	expect to see inter-rater reliability testing to be reported here, along with a description of the number of categories that required resolution via consensus. There is also insufficient detail provided around the consensus process; was this just between the two independent coders, or were others from the team involved? 5. Methods: I found the analysis section to be incomplete. In Table 1, p values are reported for differences in professions by age, gender etc. However no mention of the statistical test is included here or in the text. Nor is there an explanation of why differences were examined statistically. I would expect the dataset to be described as per Table 1, but any differences (or not should that have been the case) would not necessarily be meaningful. If anything, the significant differences in these variables suggests that they will be a confounding variable to the main outcome measures. Not only are they different professions (the independent variable), but there are fundamental differences between the two 'populations' that could account for differences in complaints. The authors do not appear to address this in their methods. 6. Findings: in addition to the previous comment, p values are reported but no other test statistic is. 7. Findings: IRRs are frequently (but curiously not always) reported without confidence intervals. I also expect to see p values and other necessary test statistics to support statements of significant differences. 8. Findings: just before the placeholder for Figure 1, it is reported that: “While psychologists had higher risk of complaints about procedures than other allied health practioners [sic], the difference was not significant...” I have two issues with this; firstly, again no test statistics or p value is provided to support this statement. Secondly, it is simply irresponsible to state that there is a higher risk of complaints when there is actually no significant difference, as the increased risk is, by definition, likely to be due to chance alone. 9. Discussion: as with the introduction, many sections of the discussion are poorly developed and do not link into the research evidence or policy base. To give an example, in ‘The meaning of the study’ section, the ‘boundaries’ sub-section provides a good overview of the surrounding evidence with five, albeit largely dated, references. Yet the ‘confidentiality’, ‘interpersonal behaviour’, ‘communication’, ‘prescribing’ and ‘regional psychologists’ sub-sections do not contain a single reference. This leads to a rather descriptive account. 10. Formatting: The standard referencing style for the journal is numbered (using an adapted Vancouver style) rather than using Harvard. Likewise, tables should be embedded within location in the text. The authors should check journal requirements before submission: https://authors.bmj.com/writing-and-formatting/formatting-your-paper/ 11. Title: I don't feel like the title accurately reflects the paper. Reference to a hotspot suggests spatial/geographical data, which this study does not include. The study also includes findings in relation to professions other than psychiatrists and psychologists.

VERSION 1 – AUTHOR RESPONSE

Thank you to Reviewer 1 for their comments in support of our manuscript.

We have incorporated Reviewer 2’s suggestions in our resubmission as follows:

1. Abstract: the results section of the abstract is missing confidence intervals and p values for the reported statistics. These are required according to the journal guidelines.

We were unable to add confidence intervals to all reported statistics in the abstract within the word limits allowed by *BMJ Open*. To work around this, we have removed IRRs relating to specific complaint issues from the abstract. We provide confidence intervals for all IRRs in the results section. If the editors wish to include this detail in the abstract, the results are as follows:

"Their risk of complaints was especially high in relation to reports (psychiatrists vs physicians IRR 9.0, 95% CI 6.7 to 12.2; psychologists vs other allied health practitioners IRR 10.9, 95% CI 7.6 to 15.9), records (psychiatrists IRR 5.6, 95% CI 4.2 to 7.6; psychologists IRR 4.7, 95% CI 3.5 to 6.3), confidentiality (psychiatrists IRR 5.0, 95% CI 3.2 to 7.8; psychologists IRR 7.9, 95% CI 5.7 to 11.2), interpersonal behaviour (psychiatrists IRR 3.2, 95% CI 2.7 to 3.9; psychologists IRR 3.6, 95% CI 3.0 to 4.5) and sexual boundary breaches (psychiatrists IRR 3.0, 95% CI 2.2 to 4.2; psychologists 2.6, 95% CI 2.0 to 3.3)."

2. Introduction: Whilst the introduction addresses differences in complaints relating to mental and physical health practitioners, it is lacking in a critical analysis of the literature surrounding complaints in healthcare and their implications for practice and policy. For instance, I would expect to see discussion surrounding the role that complaints have within the healthcare sector and for service improvement (linked to the previous comment about being 'informative'). There is a debate around whether complaints are a valid source of data for service improvement (including how they compare to other data sources), or whether they are treated as a litigation-avoidance exercise.

We have expanded the Introduction and added additional references in response to this feedback.

3. Introduction: In the hypothesis, the use of the word 'informative' is fairly meaningless. There needs to be an explanation as to why it could be informative, and to whom it would be informative for. This ties into the previous issue.

We have added an explanation in the Introduction as to why this would be informative and to whom.

4. Methods: two reviewers independently coded 149 categories into 16 complaint issues, with differences resolved by consensus. I expect to see inter-rater reliability testing to be reported here, along with a description of the number of categories that required resolution via consensus. There is also insufficient detail provided around the consensus process; was this just between the two independent coders, or were others from the team involved?

Any differences were resolved by consensus between the two reviewers only, with review by co-author DS. Inter-rater reliability testing was not conducted. We have made a small amendment to the Methods section to provide greater detail.

5. Methods: I found the analysis section to be incomplete. In Table 1, p values are reported for differences in professions by age, gender etc. However, no mention of the statistical test is included here or in the text. Nor is there an explanation of why differences were examined statistically. I would expect the dataset to be described as per Table 1, but any differences (or not should that have been the case) would not necessarily be meaningful. If anything, the significant differences in these variables suggests that they will be a confounding variable to the main outcome measures. Not only are they different professions (the independent variable), but there are fundamental differences between the two 'populations' that could account for differences in complaints. The authors do not appear to address this in their methods.

We have added additional text making it clear that univariate differences were examined using chi-squared tests. Later in the paper we present the results of multi-variate analyses, which adjust for age, sex and practice location. This adjustment addresses the risk that Reviewer 2 describes. We agree that without this adjustment, there would otherwise be a risk of confounding.

6. Findings: in addition to the previous comment, p values are reported but no other test statistic is.

All information about the test is contained in the p-value. We could report chi-squared values but these would not add any new information (since the p-value is directly derived from this). Doing so would also clutter the table.

7. Findings: IRRs are frequently (but curiously not always) reported without confidence intervals. I also expect to see p values and other necessary test statistics to support statements of significant differences.

We have added confidence intervals to IRRs throughout the manuscript. It is our view that providing both p-values and confidence intervals for the incidence rate ratios throughout the text would add many words without adding value for the reader. If the null value (IRR=1.0) is not contained within the 95% confidence interval, then the probability that the null is the true value is less than 5% or $p < 0.05$.

8. Findings: just before the placeholder for Figure 1, it is reported that: "While psychologists had higher risk of complaints about procedures than other allied health practitioners, the difference was not significant..." I have two issues with this; firstly, again no test statistics or p value is provided to support this statement. Secondly, it is simply irresponsible to state that there is a higher risk of complaints when there is actually no significant difference, as the increased risk is, by definition, likely to be due to chance alone.

The reviewer is correct that this finding is not statistically-significant. We have deleted our comments about procedures by psychologists.

9. Discussion: as with the introduction, many sections of the discussion are poorly developed and do not link into the research evidence or policy base. To give an example, in 'The meaning of the study' section, the 'boundaries' sub-section provides a good overview of the surrounding evidence with five, albeit largely dated, references. Yet the 'confidentiality', 'interpersonal behaviour', 'communication', 'prescribing' and 'regional psychologists' sub-sections do not contain a single reference. This leads to a rather descriptive account.

This section has been revised and additional references added.

10. Formatting: The standard referencing style for the journal is numbered (using an adapted Vancouver style) rather than using Harvard. Likewise, tables should be embedded within location in the text.

The reference style has been updated to adapted Vancouver style and the tables have been moved to within the text.

11. Title: I don't feel like the title accurately reflects the paper. Reference to a hotspot suggests spatial/geographical data, which this study does not include. The study also includes findings in relation to professions other than psychiatrists and psychologists.

We have revised the title, removing the word 'hotspots' and noting that the study includes other practitioners.